

<b>REPORT DOCUMENTATION PAGE</b>			Form Approved OMB NO. 0704-0188		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 01-06-2015		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 1-Aug-2014 - 30-Apr-2015	
4. TITLE AND SUBTITLE Final Report: Determining Dynamical Path Distributions using Maximum Relative Entropy			5a. CONTRACT NUMBER W911NF-14-1-0452		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER 611102		
6. AUTHORS Adom Giffin			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES Clarkson University 8 Clarkson Avenue CU Box 5630 Potsdam, NY 13676 -1401			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 65967-EG-II.1		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT The scientific objective was to extend MaxCal to include microscopic (data) information by deriving the explicit, general, theoretical distributions that includes both non-equilibrium, path trajectory information (as done in MaxCal) and direct, observational microscopic information. MaxCal is just The Principle of Maximum Entropy (MaxEnt) where constraints are changing in time. This simply amounts to an additional step to summarize the microstate (from MaxEnt) as it changes. <del>The MaxEnt algorithm and therefore, MaxCal, are only applicable for information in the form of expectation</del>					
15. SUBJECT TERMS Maximum Entropy, Bayes, inference, ergodic, non-equilibrium, dynamics					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Adom Giffin
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 315-268-2386



## Report Title

Final Report: Determining Dynamical Path Distributions using Maximum Relative Entropy

### ABSTRACT

The scientific objective was to extend MaxCal to include microscopic (data) information by deriving the explicit, general, theoretical distributions that includes both non-equilibrium, path trajectory information (as done in MaxCal) and direct, observational microscopic information.

MaxCal is just The Principle of Maximum Entropy (MaxEnt) where constraints are changing in time. This simply amounts to an additional step to summarize the microstate (from MaxEnt) as it changes.

The MaxEnt algorithm and therefore, MaxCal, are only applicable for information in the form of expectation values. Microscopic information does not have a place in it. However, macroscopic information is in essence a summary of microscopic information. Therefore, it makes logical sense that microscopic information would shape these distributions.

To address the main objective of the project, the next task was to show that MaxCal is a special case of MrE. MaxCal is a special case of MaxEnt and MaxEnt is a special case of MrE. Therefore, MaxCal is a special case of MrE. However, the specific objective was to determine the general theoretical distributions that include both non-equilibrium path information and observational information. This is simply an application and I include several examples to illustrate this application.

---

**Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:**

**(a) Papers published in peer-reviewed journals (N/A for none)**

Received

Paper

**TOTAL:**

**Number of Papers published in peer-reviewed journals:**

---

**(b) Papers published in non-peer-reviewed journals (N/A for none)**

Received

Paper

**TOTAL:**

**Number of Papers published in non peer-reviewed journals:**

---

**(c) Presentations**

Number of Presentations: 0.00

---

**Non Peer-Reviewed Conference Proceeding publications (other than abstracts):**

Received      Paper

**TOTAL:**

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

---

**Peer-Reviewed Conference Proceeding publications (other than abstracts):**

Received      Paper

**TOTAL:**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

---

**(d) Manuscripts**

Received      Paper

**TOTAL:**

Number of Manuscripts:

Books

Received      Book

TOTAL:

Received      Book Chapter

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	Discipline
Kushani De silva	1.00	
<b>FTE Equivalent:</b>	<b>1.00</b>	
<b>Total Number:</b>	<b>1</b>	

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
<b>FTE Equivalent:</b>	
<b>Total Number:</b>	

---

### Names of Faculty Supported

NAME

PERCENT SUPPORTED

**FTE Equivalent:**

**Total Number:**

---

### Names of Under Graduate students supported

NAME

PERCENT SUPPORTED

**FTE Equivalent:**

**Total Number:**

### Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: ..... 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense ..... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: ..... 0.00

---

### Names of Personnel receiving masters degrees

NAME

**Total Number:**

---

### Names of personnel receiving PHDs

NAME

**Total Number:**

---

### Names of other research staff

NAME

PERCENT SUPPORTED

**FTE Equivalent:**

**Total Number:**

---

**Sub Contractors (DD882)**

**Inventions (DD882)**

**Scientific Progress**

**Technology Transfer**

Determining Dynamical Path Distributions  
using Maximum Relative Entropy

ARO grant no. W911NF-14-1-0452

Final Report

Clarkson University

Adom Giffin P.I.

May 31, 2015



## Summary of project

The scientific objective was to extend MaxCal [1] to include microscopic (data) information by deriving the explicit, general, theoretical distributions that include both non-equilibrium, path trajectory information (as done in MaxCal) and direct, observational microscopic information.

Toward this end, the first task was to reproduce MaxCal and understand the logic and motivation behind it. MaxCal is just The Principle of Maximum Entropy (MaxEnt [2]) where constraints are changing in time. This simply amounts to an additional step to summarize the microstate (from MaxEnt) as it changes. This does not need to even be a temporal change; however MaxCal was conceived originally with constraints changing in time. However, the key point is that the main algorithm is unchanged. In other words, it is still "MaxEnt" and as Jaynes himself said, "...it applies equally well to any physical quantity whatsoever." [3]. I would add that it applies equally well to any information in the form of an expectation value (macroscopic information) whatsoever. It is simply a different application of the MaxEnt algorithm. Jaynes renamed it "MaxCal" to simply highlight the emphasis on the flux as opposed to the state. Therefore, I assert that MaxCal is a special case of MaxEnt, and not the other way around.

As is stated in the previous paragraph, the MaxEnt algorithm and therefore, MaxCal, are only applicable for information in the form of expectation values. Microscopic information does not have a place in it. However, macroscopic information is in essence a summary of microscopic information. Indeed, the probability distributions of both Gibbs and Boltzmann describe the 'microstates' of the system. Therefore, it makes logical sense that microscopic information would shape these distributions. Maximum relative Entropy (MrE) [4] was shown to be a generalized algorithm that includes both MaxEnt and Bayes Rule (which handles microscopic information such as data), as special cases.

To address the main objective of the project, the next task was to show that MaxCal is a special case of MrE. This step is trivial once one understands the preceding comments; MaxCal is a special case of MaxEnt and MaxEnt is a special case of MrE. Therefore, MaxCal is a special case of MrE. However, although this was the spirit of the project, the specific objective was to determine the general theoretical distributions that include both non-equilibrium path information and observational information. This is simply an application and I include several examples to illustrate this application.

Finally, since someone will undoubtedly claim a version of MrE that includes time based constraints is 'new' in the sense that MaxCal is "new". I will name this special case of MrE, MrE(t).

## Detailed progress and results

### Results of this program can be broken into four areas

- Examine the MaxCal algorithm and motivation for it
- Derive general MrE application example of non-equilibrium, path trajectory information, MrE(t)
- Illustrate MrE(t) with several explicit examples
- Current and future directions

### MaxCal algorithm

It needs to be stated over and over until it is well understood that MaxEnt is a method of *inference*. It applies equally well to any information in the form of an expectation value (macroscopic information) whatsoever. Therefore, there is no notion of equilibrium or ergodicity built into it.

## Background

Boltzmann assumed many things when creating his "entropy" or better, his distribution of microstates that a single particle to be in. For example, he needed to create the famous ergodic theorem so that he could justify each microstate being equally likely [5]. This allowed him to use a multiplicity to describe the number of particles in a particular macrostate. He also assumed that the particles did not interact. Although he did not state it explicitly, by action, he also assumed the only information he had was the total energy of the system. Using these pieces of information and assumptions in the MaxEnt algorithm, one obtains the microcanonical distribution of microstates,  $f_i$ . It is valid only at equilibrium because the assumptions used to create it are only valid at equilibrium.

$$f_i = \frac{1}{Z} e^{-\beta \epsilon_i} , \quad (1)$$

where,  $Z$ , is the partition physics and  $\epsilon_i$  is an energy state and  $\beta$  is the Lagrange multiplier that turns out to be inversely proportional to the temperature. Following this yields his entropy,

$$S_B = -N \sum_{i=1}^m \frac{n_i}{N} \log \frac{n_i}{N} = -N \sum_{i=1}^m f_i \log f_i , \quad (2)$$

where  $N$  is the number of particles. For the continuous case in momentum space,

$$S_B = -N \int d^3x d^3p f(x, p) \log f(x, p) . \quad (3)$$

Gibbs took a different approach [6]. He created a variational principle to determine the distribution of the system *at equilibrium* as was his intention. However the variational principle itself has nothing to do with an equilibrium state. It is this variational principle that is the true engine behind MaxEnt. Using this and **not** assuming collisionless particles led to his version of the equilibrium microstate. This led to his distribution,

$$f_i = \frac{1}{Z} e^{-\beta E_i} , \quad (4)$$

which is a state of the *entire* system of particles with  $E_i$  as the energy function. This can clearly be seen in the entropy form,

$$S[f] = - \int d^{3N}x d^{3N}p f(x, p) \log f(x, p) . \quad (5)$$

This is the basis of traditional statistical mechanics which is defined at equilibrium. This has been experimentally verified many times over. However, the key insight into understanding this result is that it is also valid for systems *far* from equilibrium. Just because Gibbs only used information that is very valid at equilibrium, such as the average energy, does not mean that this is any less valid for systems far from equilibrium. However, if this was used to predict the state of such a system, it would give terrible results. Maybe the results would be so bad that a random selections provides the same results. In inference, we attribute this to "noise" in the system and that the noise overwhelms the predictive ability of the model. Thus, while it may be a terrible model for a non-equilibrium system, it does not mean it is "wrong". It simple is not enough information to describe the state to the necessary precision.

## MaxEnt and MaxCal

The MaxEnt algorithm can be defined in the following way. The correct entropic form is,

$$S[p] = - \sum_i p_i \log p_i , \quad (6)$$

where  $p_i$  is the probability of microstate,  $i$ , with normalization constraint,

$$\sum_i p_i = 1 , \quad (7)$$

and general macroscopic constraint,

$$\sum_i p_i f^k(x_i) = F^k , \quad (8)$$

where,  $f^k(x_i)$  is the  $k^{th}$  function that is constraining the probability and  $F^k$  is the value of the expectation value to be constraining the posterior. Maximization of the entropy with constraints over the index "k" is,

$$S[p] = S[p] - \alpha \left( \sum_i p_i - 1 \right) - \sum_k \left( \lambda_k \left( \sum_i p_i f^k(x_i) - F^k \right) \right), \quad (9)$$

which yields,

$$p_i = \frac{1}{Z} e^{-\sum_k (\lambda_k f^k(x_i))}, \quad (10)$$

where  $Z$  is the typical partition function.

The MaxCal algorithm can be defined in the following way from above. Let  $k$  be a temporal index. We can rewrite the constraint (8) as,

$$\sum_i p_i \left( \sum_k f^k(x_i) \right) = \sum_i p_i f(x_i) = F. \quad (11)$$

Maximization of the entropy is,

$$S[p] = S[p] - \alpha \left( \sum_i p_i - 1 \right) - \sum_i \left[ p_i \left( \sum_k \lambda(k) f^k(x_i) - F \right) \right], \quad (12)$$

which yields,

$$p_i = \frac{1}{Z} e^{-\left( \sum_k (\lambda(k) f^k(x_i)) \right)}, \quad (13)$$

where the Lagrange multiplier,  $\lambda(k)$  enforces each discrete "time" step and now  $p_i$  is the probability of the *microtrajectory* with index "i". A common objection is that if one time step is at "equilibrium", how can it have *another* time step after? The answer follows from the argument presented above; it was never stated that the system was ever at equilibrium. The MaxEnt algorithm is silent on this issue. Is this function valid for systems far from equilibrium? It depends on the precision needed. For greater precision, more information may be needed, i.e. more constraints. Note, that if the relative entropy was used the result would be,

$$p_i = q_i \frac{1}{Z} e^{-\left( \sum_k (\lambda(k) f^k(x_i)) \right)}, \quad (14)$$

where  $q_i$  is the "prior".

## MrE(t)

Here the general MrE application example of non-equilibrium, path trajectory information, MrE(t), is derived.

## MrE

Maximum relative Entropy can be written as the Maximum relative Entropy (MrE) [4] method is designed to update from a prior to a posterior distribution on the basis of three pieces of information: prior information about  $\theta$  (the prior), the known relationship between  $x$  and  $\theta$  (the model), and the observed values of the data  $x \in \mathcal{X}$ . Since we are concerned with both  $x$  and  $\theta$ , the relevant space is neither  $\mathcal{X}$  nor  $\Theta$  but the product  $\mathcal{X} \times \Theta$  and our attention must be focused on the joint distribution  $P(x, \theta)$ . The selected joint posterior  $P_{\text{new}}(x, \theta)$  is that which maximizes the entropy<sup>1</sup>,

$$S[P, P_{\text{old}}] = - \int P(x, \theta) \log \frac{P(x, \theta)}{P_{\text{old}}(x, \theta)} dx d\theta, \quad (15)$$

---

<sup>1</sup>In the MrE terminology, we "maximize" the negative relative entropy,  $S$  so that  $S \leq 0$ . This is the same as minimizing the relative entropy.

subject to the appropriate constraints (parameters can be discrete as well).  $P_{\text{old}}(x, \theta)$  contains our prior information which we call the *joint prior*. To be explicit,

$$P_{\text{old}}(x, \theta) = P_{\text{old}}(x) P_{\text{old}}(\theta|x) , \quad (16)$$

where  $P_{\text{old}}(x)$  is the traditional Bayesian prior and  $P_{\text{old}}(\theta|x)$  is the likelihood. It is important to note that they *both* contain prior information. The Bayesian prior is defined as containing prior information. However, the likelihood is not traditionally thought of in terms of prior information. Of course it is reasonable to see it as such because the likelihood represents the model (the relationship between  $\theta$  and  $x$ ) that has already been established. Thus we consider both pieces, the Bayesian prior and the likelihood to be *prior* information. It should be noted that Shore and Johnson [7] never make this connection.

The new information is the *observed data*,  $\theta'$ , which in the MrE framework must be expressed in the form of a constraint on the allowed posteriors. The family of posteriors that reflects the fact that  $\theta$  is now known to be  $\theta'$  is such that

$$P(x) = \int P(x, \theta) dx = \delta(\theta - \theta') , \quad (17)$$

where  $\delta(\theta - \theta')$  is the Dirac delta function (or a Kronecker delta for the discrete case). This amounts to an *infinite* number of constraints: there is one constraint on  $P(x, \theta)$  for each value of the variable  $x$  and each constraint will require its own Lagrange multiplier  $\lambda(x)$ . Furthermore, we impose the usual normalization constraint,

$$\int P(x, \theta) dx d\theta = 1 , \quad (18)$$

and include additional information about  $\theta$  in the form of a constraint on the expected value of some function  $f(\theta)$ ,

$$\int P(x, \theta) f(x) dx d\theta = \langle f(x) \rangle = F . \quad (19)$$

The final step is to marginalize the posterior,  $P_{\text{new}}(x, \theta)$  over  $x$  to get our updated probability,

$$P_{\text{new}}(\theta) = P_{\text{old}}(x, \theta') \frac{e^{\beta f(\theta)}}{\zeta(\theta', \beta)} \quad (20)$$

and  $\zeta(\theta', \beta)$  is the partition function and  $\beta$  is the Lagrange multiplier.

## MrE(t)

In the above example, no mention of equilibrium was ever made. Now a temporal function is explicitly employed in the constraint,

$$\int P(x(t), \theta(t)) \left( \int f(x(t), t) dt \right) dx(t) d\theta(t) = \langle f(x(t)) \rangle = F . \quad (21)$$

and therefore the final distribution would be,

$$P_{\text{new}}(x(t)) = P_{\text{old}}(x(t), \theta(t)') \frac{e^{\int \beta(t) f(x(t), t) dt}}{\zeta(\theta(t)', \beta(t))} \quad (22)$$

where  $\beta(t)$  is an infinite amount of Lagrange multipliers for the time constraints. To provide a direct comparison to the discrete MaxCal above, we have,

$$p_i = q_{ij'} \frac{1}{Z} e^{-\left( \sum_k \lambda^{(k)} f^k(x_i) \right)} , \quad (23)$$

where the index "j" is for the observable constraint which in this case is not a Dirac delta function but a Kronecker delta function,  $\delta_{jj'}$ ,  $q_{ij'}$  is the discrete Bayesian prior and the partition function,  $Z$ , is a function of the observable index "j" as well.

## Examples

Here I provide a few examples to illustrate MrE(t).

## General particle in motion

Let a particle's path in time be defined by a continuous function,  $f(x, \dot{x}; t)$ . The constraint that would be employed in MaxCal is,

$$\int P(x(t)) \left( \int f(x, \dot{x}; t) dt \right) dx(t) = \langle f(x, \dot{x}; t) \rangle, \quad (24)$$

with distribution,

$$P(x(t)) = \frac{e^{-\int \lambda(t) f(x, \dot{x}; t) dt}}{Z(\lambda(t))}, \quad (25)$$

where  $\lambda(t)f(x, \dot{x}; t)$  is the classical Lagrangian and the integration produces the action. In contrast, MrE(t) would be,

$$P_{\text{new}}(x(t)) = P_{\text{old}}(x(t), \theta(t)') \frac{e^{-\int \lambda(t) f(x, \dot{x}; t) dt}}{\zeta(\theta(t)', \beta(t))}, \quad (26)$$

where  $\beta(t)$  is the Lagrange multiplier and  $\zeta(\theta(t)', \beta(t))$  is the partition function.

## Specific particle in motion example (1D)

Let's assume that the motion was very simple in that the particle has constant velocity,  $v$ , no potential energy acting on it and is moving in one dimension. The action over the time interval  $[0, t]$  would then be

$$x(t) = x(0) + vt \quad (27)$$

and so the distribution of the paths with MaxCal would be,

$$P(x(t)) = \frac{e^{-x(0) - vt}}{Z(\lambda(t))} \quad (28)$$

and the MrE(t) solution would be,

$$P_{\text{new}}(x(t)) = P_{\text{old}}(x(t), \theta(t)') \frac{e^{-x(0) - vt}}{\zeta(\theta(t)', \beta(t))} \quad (29)$$

where the Bayesian prior,  $P_{\text{old}}(x(t), \theta(t)')$  is some function that relates the position with some other observable. Perhaps a magnetic field. However, while the microscopic observables certainly change the shape of the distribution, the mean of the position is assumed to be known in order to determine the Lagrange multipliers. Therefore, the microscopic observables do not have any influence on the mean,  $\langle x(t) \rangle$

## Specific particle in motion example (2D\_1)

Let's assume that the motion was very simple in that the particle has constant velocity,  $v$ , no potential energy acting on it and is moving in *two* dimensions. The action over the time interval  $[0, t]$  would then be

$$x(t) = x(0) + v_x t \quad (30)$$

$$y(t) = y(0) + v_y t \quad (31)$$

and so the distribution of the paths with MaxCal would be,

$$P(x(t)) = \frac{e^{-x(0) - v_x t - y(0) + v_y t}}{Z(\lambda_x(t), \lambda_y(t))} \quad (32)$$

and the MrE(t) solution would be,

$$P_{\text{new}}(x(t), y(t)) = P_{\text{old}}(x(t), y(t), \theta_x(t)', \theta_y(t)') \frac{e^{-x(0) - v_x t - y(0) + v_y t}}{\zeta(\theta_x(t)', \theta_y(t)', \beta(t))} \quad (33)$$

where the Bayesian prior,  $P_{\text{old}}(x(t), \theta(t)')$  is some function that relates the position with some other observable.

## Specific particle in motion example (2D\_2)

Now let's assume that we do not know the means ahead of time, which is a much more plausible scenario. This is especially true if the position is not well observed in time, i.e. the sample mean,  $\bar{x}(t)$  would be a very poor estimate of the mean  $\langle x(t) \rangle$ . However, we may know some other important information regarding the velocities. Such as that the moment is conserved. This means that on average over a long time, the velocity components will be related in some way,

$$\langle x(t) \rangle = g_0 \langle y(t) \rangle \quad (34)$$

where  $g_0$  is the ratio of the average of the initial velocity components. In this case the constraint would be,

$$\int P(x(t), y(t)) \left( \int [f(x, \dot{x}; t) - g_0 f(y, \dot{y}; t)] dt \right) dx(t) dy(t) = \langle f(x, \dot{x}; t) \rangle - g_0 \langle f(y, \dot{y}; t) \rangle \quad (35)$$

Now the prior and thus the observables will influence what the mean of the system is can be.

## Flux of states

Here I examine the Ehrenfest's famous "dog-flea" model. In this discrete case, fleas either jump off a dog or stay on. This can be seen as a flux of fleas. When this is combined with other dogs who are exchanging fleas, it turns into a simplified version of many real world physical systems. However, for our example, I will limit myself at first with the flux of one dog. Further, I will describe the systems as a set of coins, where each flea is a coin and in a state of 1 for jumping off or 0 for staying on. Let the index " $i$ " denote a microstate of the system with  $N$  microstates. For example, if there were 4 fleas on the dog or 4 coins, there would be  $16 = 2^4$  microstates. We knew the average number of jumps or heads (changes),  $\langle m \rangle$ , the entropy to be maximized would be,

$$S[p] = - \sum_i p_i \log p_i - \alpha \sum_i p_i - \lambda \sum_i p_i m_i, \quad (36)$$

and after some manipulations, the distribution for the microstates, would be converted to the distribution of the jumps, where the binomial distribution would be produced,

$$p(m) = \frac{N!}{m!(N-m)!} p^m (1-p)^{(N-m)}. \quad (37)$$

We can extend this to more than a two state system, such as instead of a coin, we have a 3 sided die. Or the states are described by,  $-1, 0, 1$  that might represent two other dogs that the flea can jump too. With 4 fleas we would now have  $3^4 = 81$  microstates. In that case the result would be,

$$p(m_1, m_2) = \frac{N!}{m_1! m_2! (N - m_1 - m_2)!} p_1^{m_1} p_2^{m_2} (1 - p_1 - p_2)^{(N - m_1 - m_2)}. \quad (38)$$

This once again assumes we know both the means of the two dimensions of this 2D-simplex ( $m_3 = N - m_1 - m_2$ ). If this is not the case, and we employ a similar constrain as in the practical example, we would have,

$$S[p] = - \sum_i p_i \log p_i - \alpha \sum_i p_i - \lambda \sum_i p_i (m_{1i} - g m_{2i}) \quad (39)$$

with some prior that relates some observables to the  $m$ 's. For a more detailed example of a more complicated system, see ([8]). These examples can immediately produce new views of Fick's law of diffusion, Fourier's law of heat flow, the Newtonian viscosity law, and the mass-action laws of chemical kinetics.

## Final thoughts and future work

In [9] we adopted a consistency axiom similar to that proposed by Shore and Johnson [7]. When two systems are independent it should not matter whether the inference procedure treats them separately or jointly. The merit of such a consistency axiom is that it is very compelling: it is difficult to advocate any other alternative. Nevertheless this axiom has been criticized by Karbelkar [10] and by Uffink [11]. In their view it fails to single out the usual logarithmic entropy as the unique tool for updating. It merely restricts the form of the entropy to a one-dimensional continuum labeled by a parameter  $\eta$ . The resulting  $\eta$ -entropies are equivalent to those proposed by Renyi [12] or by Tsallis [13] in the sense that they lead to the same updated probabilities.

The main result of [9] was to go beyond the insights of Karlbekar and Uffink, and show that the consistency axiom selects a unique, universal value for the parameter  $\eta$  and this value corresponds to the usual logarithmic entropy. The advantage of our approach is that it shows precisely how it is that the other  $\eta$ -entropies are ruled out as tools for updating. This is particularly important as the two, recent, "hot" articles cited as the state of the art above [14, 15] use Shore and Johnson as their foundation. We have *already* addressed the criticism that *will* be applied to those articles.

The ideas of MaxEnt have evolved as well. MaxEnt was designed to *assign* rather than update probabilities. However, if information is given in the form of data, then the proper method for inference was Bayes theorem. The method of Maximum relative Entropy (MrE) [4] is capable of reproducing every aspect of orthodox Bayesian inference and proves the complete compatibility of Bayesian and Maximum Entropy methods. However, it also opens the door to tackling problems that could not be previously addressed by either the MaxEnt or orthodox Bayesian methods individually, such as inferring parameters from both constraints and data *simultaneously*. This fundamentally changes the inference landscape as now there is no separation of the microscopic (measurements, observations, data) and the macroscopic (constraints, moments, Hamiltonians, averages, etc.).

Besides extending the many above examples, there are many, many uses of MaxCal that can be exploited with the generalized algorithm, MrE(t). Fluids and gas dynamics are two applicaitons that would be very exciting to pursue as would biological applications such as protein folding. All of these have direct relevance to Army objectives in complex, nonlinear, dynamical systems, fluid dynamics, self-organizing systems and large scale, autonomous, dynamical networks.

## References

- [1] E. T. Jaynes, "Macroscopic Prediction", Complex Systems-Operational Approaches in Neurobiology, Physics, and Computers, H. Haken, Ed., Springer, Berlin (1985); p. 254
- [2] E. T. Jaynes, Phys. Rev. **106**, 620 and **108**, 171 (1957).
- [3] E. T. Jaynes, 'Foundations of Probability Theory and Statistical Mechanics,' in Delaware Seminar in the Foundations of Physics, M. Bunge (ed.), Springer-Verlag, Berlin, (1967); p. 77
- [4] A. Giffin and A. Caticha, "Updating Probabilities with Data and Moments", *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, eds. K. Knuth, A. Caticha, J. L. Center, A. Giffin, C. C. Rodriguez, AIP Conf. Proc. **954**, 74 (2007).
- [5] J. Uffink, "Boltzmann's Work in Statistical Physics", *Stanford Encyclopedia of Philosophy*, (2004) references L. Boltzmann, *Wiener Berichte* **76**, 373, (1877).
- [6] J. W. Gibbs, *Elementary Principles in Statistical Mechanics*, (1902); Reprinted in Collected works and commentary, (Yale University Press 1936), and (Dover Publications, Inc. 1960).
- [7] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theory **IT-26**, 26 (1980); IEEE Trans. Inf. Theory **IT-27**, 26 (1981).
- [8] A. Giffin, "Inferring Diversity: Life after Shannon", *InterJournal of Complex Systems*, 2201 (2008) (<http://arxiv.org/abs/0709.4079>).
- [9] A. Caticha and A. Giffin, "Updating Probabilities", *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, ed. A. Mohammad-Djafari, AIP Conf. Proc. **872**, 31 (2006).
- [10] S. N. Karbelkar, Pramana – J. Phys. **26**, 301 (1986).
- [11] J. Uffink, "Can the maximum entropy principle be explained as a consistency requirement?", *Studies in History and Philosophy of Modern Physics* **26B**, 223 (1995) *Studies in History and Philosophy of Modern Physics* **26B**, 223 (1995).
- [12] A. Renyi, "On measures of entropy and information," *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol 1, p. 547-461 (U. of California Press, 1961).
- [13] C. Tsallis, J. Stat. Phys. **52**, 479 (1988); "Nonextensive statistical mechanics: a brief review of its present status," online at [arXiv.org/abs/cond-mat/0205571](http://arXiv.org/abs/cond-mat/0205571).
- [14] J. Peterson, P. D. Dixit, and K. A. Dill, "A maximum entropy framework for nonexponential distributions", *PNAS* (2013).
- [15] Presse, S., et al. "Principles of maximum entropy and maximum caliber in statistical physics". *Reviews of Modern Physics*, 85, (2013).